

## CLAIMS

I Claim:

- 1 1. A method of recognizing a received phoneme using a stored plurality of
- 2 phoneme classes, each of the plurality of phoneme classes comprising class phonemes,
- 3 the method comprising:
- 4 (A) training the class phonemes, the training comprising, for each class
- 5 phoneme:
- 6 (1) determining a phoneme vector as a time-frequency representation
- 7 of the class phoneme;
- 8 (2) dividing the phoneme vector into phoneme segments;
- 9 (3) assigning each phoneme segment into a plurality of phoneme
- 10 parameters;
- 11 (4) expanding each phoneme segment and plurality of phoneme
- 12 parameters into an expanded stored-phoneme vector with expanded vector parameters;
- 13 (5) transforming the expanded stored-phoneme vector into an
- 14 orthogonal form using singular-value decomposition wherein:
- 15  $[x_1 \ x_2 \ \dots \ x_m] = [u_1 \ u_2 \ \dots \ u_m] \Lambda V^t$ , where  $x_k$  is a  $k^{\text{th}}$  acoustic vector for a corresponding
- 16 stored phoneme,  $u_k$  is the corresponding orthogonal vector and  $\Lambda$  and  $V$  are diagonal
- 17 and unitary matrices, respectively; and
- 18 (B) recognizing the received phoneme by:
- 19 (1) receiving an analog acoustic signal;
- 20 (2) converting the analog acoustic signal into a digital signal;
- 21 (3) determining a received-signal vector as a time-frequency
- 22 representation of the received digital signal;
- 23 (4) dividing the received-signal vector into received-signal segments;

24 (5) assigning each received-signal segment into a plurality of received-  
25 signal parameters;

26 (6) expanding each received-signal segment and plurality of received-  
27 signal parameters into an expanded received-signal vector;

28 (7) transforming the expanded received-signal vector into an  
29 orthogonal form using singular-value decomposition wherein:

30  $[y_k] = [z_k] \Lambda V^t$ , where  $y_k$  is a  $k^{\text{th}}$  acoustic vector for a corresponding received phoneme,  $z_k$   
31 is the corresponding orthogonal vector and  $\Lambda$  and  $V$  are diagonal and unitary matrices,  
32 respectively;

33 (8) determining a first distance associated with the orthogonal form  
34 of the expanded received-signal vector and a second distance associated respectively with  
35 each orthogonal form of the expanded stored-phoneme vectors; and

36 (9) recognizing the received phoneme according to a comparison of  
37 the first distance with the second distance.

1 2. The method of claim 1, wherein transforming the expanded stored-phoneme  
2 vector into an orthogonal form using singular-value decomposition and wherein  
3 transforming the expanded received-signal vector into an orthogonal form using singular-  
4 value decomposition conforms the stored-phoneme vector and the expanded received-  
5 signal vector into a hypersphere having a center and a radius.

1 3. The method of claim 2, wherein determining a distance associated with the  
2 orthogonal form of the expanded received-signal vector and each orthogonal form of the  
3 expanded stored-phoneme vectors further comprises:

4 comparing a distance from the center of the hypersphere of the orthogonal form  
5 of the expanded received-signal vector with a distance from the center of the  
6 hypersphere for each orthogonal form of the expanded stored-phoneme vector.

- 1 4. The method of claim 3, wherein determining a distance associated with the  
2 orthogonal form of the expanded received-signal vector and each orthogonal form of the  
3 expanded stored-phoneme vectors further comprises:  
4 determining a difference between the distance from the center of the hypersphere  
5 of the orthogonal form of the expanded received-signal vector and the distance from the  
6 center of the hypersphere for each orthogonal form of the expanded stored-phoneme  
7 vectors, wherein the expanded stored-phoneme vectors associated with m-shortest  
8 differences between the distance from the center of the hypersphere of the orthogonal  
9 form of the expanded received-signal vector and the distance from the center of the  
10 hypersphere for each orthogonal form of the expanded stored-phoneme vectors are  
11 recognized as most likely to be associated with the received phoneme.
- 1 5. The method of claim 1, wherein the orthogonal form of the expanded stored-  
2 phoneme vector and the expanded received-signal vector each have at least  
3 approximately 100 dimensions.
- 1 6. The method of claim 1, wherein each acoustic vector for a corresponding stored  
2 phoneme has a mean value removed.
- 1 7. The method of claim 6, wherein each acoustic vector for a corresponding  
2 received phoneme has a mean value removed.
- 1 8. The method of claim 1, wherein the phoneme vector determined as a time-  
2 frequency representation of the class phoneme is a representation of approximately 125  
3 msec.
- 1 9. The method of claim 8, wherein the phoneme vector is divided into  
2 approximately 25 msec phoneme segments.
- 1 10. The method of claim 9, wherein each 25 msec phoneme segment is assigned  
2 approximately 32 phoneme parameters.

- 1 11. The method of claim 10, wherein each of the approximately 25 msec phoneme  
2 segments with 32 phoneme parameters is expanded into an expanded stored-phoneme  
3 vector with approximately 160 parameters.
- 1 12. The method of claim 11, wherein the received-signal vector determined as a time-  
2 frequency representation of the received digital signal is a representation of  
3 approximately 125 msec.
- 1 13. The method of claim 11, wherein the received-signal vector is divided into  
2 approximately 25 msec received-signal segments.
- 1 14. The method of claim 13, wherein each approximately 25 msec received-signal  
2 segment is assigned approximately 32 received-signal parameters.
- 1 15. The method of claim 14, wherein each of the approximately 25 msec received-  
2 signal segments with 32 received-signal parameters is expanded into an expanded  
3 received-signal vector with approximately 160 parameters.
- 1 16. A method of recognizing speech patterns, the method using stored phonemes,  
2 the method comprising:  
3 converting each stored phoneme into n-dimensional space having a center;  
4 sampling speech patterns to obtain at least one sampled phoneme;  
5 converting each of the at least one sampled phonemes into the n-dimensional  
6 space; and  
7 comparing a distance from the center of the n-dimensional space to the sampled  
8 phoneme with a distance from the center of the n-dimensional space to each of the  
9 phonemes of the converted plurality of phonemes.
- 1 17. The method of claim 16, wherein converting the stored phonemes comprises  
2 using singular-value decomposition.
- 1 18. The method of claim 16, further comprising storing the converted phonemes  
2 before sampling speech patterns.

1 19. The method of claim 16, wherein n equals at least 100.

1 20. The method of claim 16, wherein comparing the distance from the center of the  
2 n-dimensional space to the sampled phoneme with the distance from the center of the n-  
3 dimensional space to each of the converted phonemes further comprises:  
4 determining a difference between the distance from the center of the n-dimensional  
5 space to the sampled phoneme with the distance from the center of the n-dimensional  
6 space to each of the converted phonemes.

1 21. The method of claim 20, further comprising:

2 recognizing the sampled phoneme as the stored phoneme associated with the  
3 smallest difference between the distance from the center of the n-dimensional space to  
4 the sampled phoneme with the distance from the center of the n-dimensional space to  
5 each of the converted phonemes.

1 22. The method of claim 16, wherein the n-dimensional space is hyperspherical.

1 23. The method of claim 16, wherein converting the stored plurality of phonemes  
2 into n-dimensional space having a center further comprises:

3 assigning a stored-phoneme vector having approximately 160 parameters to each  
4 stored phoneme; and

5 transforming each stored-phoneme vector into the n-dimensional space having  
6 the center, wherein a probability density of the stored phonemes in the n-dimensional  
7 space is approximately spherical.

1 24. The method of claim 23, wherein converting each of the at least one sampled  
2 phonemes into the n-dimensional space further comprises:

3 assigning a sampled-phoneme vector having approximately 160 parameters to  
4 each sampled phoneme; and

5 transforming each sampled-phoneme vector into the n-dimensional space having  
6 the center, wherein a probability density of the stored phonemes in the n-dimensional  
7 space is approximately spherical.

1 25. A method of recognizing speech using a database of stored phonemes converted  
2 into n-dimensional space, the method comprising:

3 receiving a received phoneme;

4 converting the received phoneme to n-dimensional space;

5 comparing the received phoneme to each of the stored phonemes in n-  
6 dimensional space; and

7 recognizing the received phoneme according the comparison of the received  
8 phoneme to each of the stored phonemes.

1 26. The method of recognizing speech according to claim 25, wherein comparing the  
2 received phoneme to each of the stored phonemes in n-dimensional space further  
3 comprises:

4 comparing a first distance from a center of the n-dimensional space to a first  
5 point associated with the received phoneme with a second distance from the center of  
6 the n-dimensional space to a second point associated in turn with each of the stored  
7 phonemes.

1 27. The method of claim 26, wherein "n" is at least approximately 100.

1 28. The method of claim 26, wherein comparing the first distance with the second  
2 distance for each of the stored phonemes further comprises:

3 determining a difference between the first distance and the second distance for  
4 each stored phoneme.

1 29. The method of claim 28, wherein recognizing the received phoneme according  
2 the comparison of the received phoneme to each of the stored phonemes further  
3 comprises:

4 recognizing the received phoneme according to the stored phoneme associated  
5 with the smallest difference between the first distance and the second distance.

1 30. A system for recognizing phonemes, the system using a database of stored  
2 phonemes for comparison with received phonemes, the stored phonemes having been  
3 converted into n-dimensional space, the system comprising:

4 a recording element that receives a phoneme;

5 a computer that converts the received phoneme into n-dimensional space,  
6 wherein the computer compares in the n-dimensional space the received phoneme with  
7 each phoneme in the database of stored phonemes.

1 31. The system of claim 30, wherein the computer recognizes the received phoneme  
2 using the comparison in the n-dimensional space of the received phoneme with each  
3 phoneme in the database of stored phonemes.

1 32. The system of claim 31, wherein the computer compares the received phoneme  
2 with each phoneme in the database of stored phonemes by comparing a first distance  
3 from a center of the n-dimensional space to a first point associated with the received  
4 phoneme with a second distance from the center of the n-dimensional space to a second  
5 point associated with each respective stored phoneme from the database of stored  
6 phonemes.

1 33. The system of claim 32, wherein the computer recognizes the received phoneme  
2 by determining a difference between the first distance and the second distance.

1 34. The system of claim 33, wherein the computer recognizes the received phoneme  
2 as associated with a stored phoneme corresponding to a shortest distance between the  
3 first distance and the second distance.

1 35. A medium storing a program for instructing a computer device to recognize a  
2 received speech signal using a database of stored phonemes converted into n-

3 dimensional space, the program comprising instructing the computer device to perform  
4 the following steps:

5 receiving a received phoneme;

6 converting the received phoneme to n-dimensional space;

7 comparing the received phoneme to each of the stored phonemes in n-  
8 dimensional space; and

9 recognizing the received phoneme according the comparison of the received  
10 phoneme to each of the stored phonemes.

1 36. A medium storing a program for instructing a computer device to recognize a  
2 received speech signal using a database of stored phonemes converted into n-  
3 dimensional space, the database of stored phonemes formed by training the stored  
4 phonemes according to the following steps:

5 (1) determining a phoneme vector as a time-frequency representation of the  
6 stored phoneme;

7 (2) dividing the phoneme vector into phoneme segments;

8 (3) assigning each phoneme segment into a plurality of phoneme parameters;

9 (4) expanding each phoneme segment and plurality of phoneme parameters  
10 into an expanded stored-phoneme vector with expanded vector parameters;

11 (5) transforming the expanded stored-phoneme vector into an orthogonal  
12 form using singular-value decomposition wherein:

13  $[x_1 \ x_2 \ \dots \ x_m] = [u_1 \ u_2 \ \dots \ u_m] \Lambda V^t$ , where  $x_k$  is a  $k^{\text{th}}$  acoustic vector for a corresponding  
14 stored phoneme,  $u_k$  is the corresponding orthogonal vector and  $\Lambda$  and  $V$  are diagonal  
15 and unitary matrices, respectively, the program stored on the medium instructing the  
16 computer device to perform the following steps:

17 (1) receiving an analog acoustic signal;

18 (2) converting the analog acoustic signal into a digital signal;



- 19           (3)     determining a received-signal vector as a time-frequency representation of  
20 the received digital signal;
- 21           (4)     dividing the received-signal vector into received-signal segments;
- 22           (5)     assigning each received-signal segment into a plurality of received-signal  
23 parameters;
- 24           (6)     expanding each received-signal segment and plurality of received-signal  
25 parameters into an expanded received-signal vector;
- 26           (7)     transforming the expanded received-signal vector into an orthogonal  
27 form using singular-value decomposition wherein:  
28  $[y_k] = [z_k] \Lambda V^t$ , where  $y_k$  is a  $k^{\text{th}}$  acoustic vector for a corresponding received phoneme,  $z_k$   
29 is the corresponding orthogonal vector and  $\Lambda$  and  $V$  are diagonal and unitary matrices,  
30 respectively;
- 31           (8)     determining a first distance associated with the orthogonal form of the  
32 expanded received-signal vector and a second distance associated respectively with each  
33 orthogonal form of the expanded stored-phoneme vectors; and
- 34           (9)     recognizing the received phoneme according to a comparison of the first  
35 distance with the second distance.